DOCUMENT RESUME

ED 308 244 TM 013 664

AUTHOR Kelderman, Henk
TITLE Common Item Equating Using the Loglinear Rasch Model. Project Psychometric Aspects of Item Banking No. 12. Research Report 86-9.
INSTITUTION Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE 86
NOTE 34p.
AVAILABLE FROM Mediatheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Equated Scores; Equations (Mathematics); Estimation (Mathematics); Foreign Countries; Latent Trait Theory; Mathematical Models; Models; Secondary Education; Secondary School Students; Solid Geometry; Testing Problems; *Test Items
IDENTIFIERS Common Item Effect; *Item Equating; Item Parameters; Log Linear Models; Netherlands; *Rasch Model

ABSTRACT

A method is proposed to equate different sets of items administered to different groups of individuals using the Rasch model. A Rasch equating model was formulated to describe one common Rasch scale in different groups with different but overlapping sets of items. The item parameters can then be estimated simultaneously, avoiding different parameter estimates of common items in different groups. The model can be tested globally to test the hypothesis of one common Rasch scale, and the goodness of link can be tested. The method is based on the quasi-loglinear Rasch model. The proposed model is illustrated on real data by equating a set of 10 solid geometry items developed and administered by the Netherlands Central Institute for Test Development to 202 students in secondary education. Five tables present the model and data from the solid geometry study. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED308244

Common Item Equating Using the Loglinear Rasch Model

Research Report 86-9

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. NELISSEN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

H. Kelderman

Division of Educational Measurement and Data Analysis

University of Twente

TM013664

Department of Education

Colophon

Typing : L. Padberg
Cover design : M. Driessen, AV-section, University of Twente
Printed by : Central Reproduction Department, University of Twente

Common Item Equating Using the Loglinear Rasch Model

Henk Kelderman

Abstract

In this paper a method is proposed to equate different sets of items administered to different groups of individuals using the Rasch model. A Rasch equating model is formulated that describes one common Rasch scale in different groups with different but overlapping sets of items. The item parameters can then be estimated simultaneously, avoiding different parameter estimates of common items in different groups. The model can be tested globally to test the hypothesis of one common Rasch scale and the goodness of link can be tested. The method is based on the quasi-loglinear Rasch model.

Keywords: test equating, latent-trait theory, Rasch model, quasi-loglinear model, incomplete contingency table, chi-square test

Common Item Equating Using the Loglinear Rasch Model

In many practical testing situations it is desirable to be able to obtain the same measurement via different instruments. For example in individualized instruction, where each student works in his own pace, it is necessary to administer different sets of items at different occasions to obtain a score that can be used to make a pass fail decision. In this case it would not be fair if a student's score could be expected to be markedly higher on one test than on another test. As another example consider the situation where a population of quite heterogeneous ability must be tested, e.g. in a longitudinal study. A number of tests of different difficulty levels must than be constructed to measure the ability of students at different levels of ability (Slinde & Linn, 1977). Also in this case it would be desirable to obtain scores on each of the tests that are comparable between all test forms.

To obtain the same measurement from the different tests, they must measure the same trait and the test scores must be measured on the same scale. In that case the scores of the tests are equated.

The traditional way to equate tests is to administer them to the same population and to transform the number-right scores of the tests to the corresponding percentile ranks (Angoff, 1971). If the tests measure the same trait this would ensure that a certain individual would have approximately the same score on both tests.

If however the tests do not measure the same trait, the individuals are not ranked the same by both tests so that they do

not have the same expected score on both tests. A way to circumvent this problem is to see that both tests conform to a common item-response models for equating (Lord, 1977). Item-response-theory equating is performed under the assumption of a common latent trait in each test. If all items under consideration fit the item-response model, all subsets of items also measure the same trait and the latent ability scores are automatically equated (Wright, 1977).

A simple and statistically attractive item-response-theory model for equating is the Rasch (1980) model (Divgi, 1981; Guskey, 1981; Gustafsson, 1979; Loyd & Hoover, 1980; Master, 1985; Renz & Bashaw, 1975; Slinde & Linn 1978, 1979; and Wright, 1977).

The Rasch model describes the probability $P(X_j=x_j|\alpha)$ that an individual with parameter α gives a response x_j to item j (j=1,...,k), where X_j can take values $x_j = 0,1$ for a wrong (0) or a right (1) response:

$$(1) \qquad P(X_j=x_j|\alpha) = \exp(x_j(\alpha-\delta_j)) \; / \; (1+\exp(\alpha-\delta_j)) \; ,$$

where δ_j (j=1,...,k) is a single item parameter describing the difficulty of item j. Making the usual assumption of local independence (Lord and Novick 1968, p. 538; Kelderman, 1984) of the item responses for a given value of α and conditioning on the number right score taking the value t, the joint probability of the item responses $X_1,...,X_k$ for a given score T=t becomes (Rasch 1966):

$$(2) \qquad P(X_1=x_1,\ldots,X_k=x_k \mid T=t)$$

$$= \exp(-x_1\delta_1\ldots-x_k\delta_k)/\Big(\sum_{x_1} \ldots \sum_{x_k} \exp(-x_1\delta_1\ldots-x_k\delta_k)\Big)$$
$$t = x_1+\ldots+x_k$$

In the Rasch model there is an indeterminacy in the scale of the item parameters. Adding a constant c to the item parameters and substracting c from the person parameter does not change model (1) and likewise in the conditional model (2) adding a constant c to each of the item parameters is equivalent to multiplying both numerator and divisor with exp $\{- t_i c\}$. The indeterminacy can be removed by fixing one item parameter to a predetermined value.

To equate different sets of items measuring the same trait, the scales must be fixed in such a way that the same person would obtain the same ability parameter regardless of the set of items (s)he would have made. This is only the case if the difficulties of the items in the different sets are on the same scale. One way to achieve this is by using one or more common items in the different sets (Wright, 1977). If the scale in the first set of items is fixed in some convenient way, the scale in the second set can now be fixed by setting the values of the common items in the second set equal to the values of the corresponding items in the set. If there are more than two sets of items, the scales can be equated if each set is directly or indirectly connected to all other sets by common items.

If the sets have more than one item in common, the usual

procedure to equate the different sets of items using the Rasch model (Wright, 1977) is to estimate the item parameters of the first set, normalize the parameters of that set, and then estimate the item parameters of the second set separately. The origin of the second scale is than transformed in such a way that the mean of the common itemparameters in the second set is the same as the mean of the common item parameter in the first set. If there are more than two sets, this procedure can be repeated for each new set. The equality of the estimates of the common item parameters in both tests can then be tested statistically by a chi-square test for goodness of link.

In this paper an alternative method of equating is proposed where all item parameters of the different sets of items are estimated simultaneously. The common item parameters are estimated only once so that, unlike in Wright's procedure, the same items have the same parameters estimates in different equating groups. If the common item sets contain more than one item, the goodness of fit of the common item links can be tested by a likelihood ratio test. To test whether the different sets of items belong to one and the same Rasch scale an overall goodness-of-fit test of the Rasch equating model is given.

The method is based on the quasi-loglinear models for incomplete contingency tables (Fienberg, 1972; Bishop, Fienberg, & Holland, 1975). Kelderman (1984) formulated the Conditional Rasch model (2) as a quasi-loglinear model where the item responses are quasi independent in the incomplete score x item 1 x item 2 x ... x item

k contingency table.

In the sequel a score x item 1 x item 2 x ... x item k contingency table is described for one equating group. A general quasi-loglinear model for this table is described and it is shown that the conditonal Rasch model is a special case of this model. Then a contingency tabel for several equating groups is described and a quasi-loglinear model is formulated for simultaneous conditional Rasch models in the different equating groups. The model is illustrated on real data by equating a set of solid geometry items.

A Quasi Loglinear Rasch Model

Let $f_{tx_1 \ldots x_k}$ be the number of individuals with number right score T=t (t=0,1,...,k) and item scores $X_1=x_1,\ldots,X_k=x_k$ where x_j = 1 if item j (j=1,...,k) is answered correctly and x_j = 0 if item j is answered incorrectly. Since it is logically impossible to have a test score that is unequal to the number of correct item responses (excluding counting errors) the counts $f_{tx_1 \ldots x_k}$ are zero for $t \neq \sum_i x_i$. Table 1 shows the score x item 1 x ... x item 3

table 1

contingency table for one equating group. Dashes denote cells that are logically or structurally zero cell. Contingency tables with

structurally zero cells are called incomplete contingency tables. Fienberg (1972; see also Bishop, Fienberg & Holland, 1975) presents a general theory for the statistical analysis of incomplete multi-way contingency tables by quasi-loglinear models.

Let $m_{tx_1 \ldots x_k}$ be the expected counts for the table under some model. If $t \neq x_1 + \ldots + x_k$ the expected counts are again structurally zero. If $t = x_1 + \ldots + x_k$, the expected counts are not structurally zero and these counts are explained by a quasi-loglinear model.

A restrictive quasi-loglinear model for the table is

$$(3) \qquad \ln m_{tx_1 \ldots x_k} = u + u_1(t) + u_2(x_1) + \ldots + u_{k+1}(x_k)$$

with the constraints

$$(4) \qquad u_1(+) = \ldots = u_{k+1}(+) = 0$$

where a plus sign indicates summation. The u-terms describe the effects of the score t and item responses $x_1, \ldots, x_k$ on the logarithm of the cell counts.

If the score is taken as a fixed variable and the item responses are considered as random variables, model (3) is equivalent to the conditional Rasch model (2). In that case $m_{tx_1 \ldots x_k}$ is the conditional expected frequency of the response $X_1=x_1, \ldots, X_k=x_k$ for given score t. The conditional probability of response $X_1=x_1, \ldots, X_k=x_k$ given t can then be obtained from (3) by

$$(5)\qquad P(X_1=x_1,\ldots,X_k=x_k \mid T=t) = m_{tx_1\ldots x_k} / \sum_{\substack{x_1 \\ x_1+\ldots+x_k=t}} \cdots \sum_{x_k} m_{tx_1\ldots x_k},$$

$$= \exp(u_3(x_1)+\ldots+u_{k+2}(x_k)) \; / \; \sum_{\substack{x_1 \\ x_1+\ldots+x_k=t}} \cdots \sum_{x_k} \exp(u_3(x_1)+\ldots+u_{k+2}(x_k))$$

Except for a reparametrization, this model is equivalent to model (2). In model (2) the effect $-x_j\delta_j$ of a response $X_j=x_j$ on item j is $-\delta_j$ for a correct response ($X_j=1$) and zero for an incorrect response ($X_j=0$), whereas in model (3) the effect of a correct response is $u_{j+1}(1)$ and the effect of an incorrect response is $u_{j+1}(0)$, where $u_{j+1}(0) = -u_{j+1}(1)$ by the constraints (4). Model (5) can be parametrized in the same way as model (2) if $u_{j+1}(1)$ is added to each parameter $u_{j+1}(x_j)$ so that the new parameter $u_{j+1}(x_j) + u_{j+1}(1)$ becomes $u_{j+1}(0) + u_{j+1}(1) = 0$ with an incorrect response and $2u_{j+1}(1)$ with a correct response. This can be done by multyplying both numerator and denominator by

$$\exp(u_3(1)+\ldots+u_{k+1}(1)),$$

so that model (5) becomes model (2) with

$$-x_j\delta_j = u_{j+1}(x_j) + u_{j+1}(1) = x_j(2u_{j+1}(1)),$$

for all j=1,...,k; i.e. $\delta_j = 2u_{j+1}(0)$. This shows that the Rasch model is equivalent to the quasi-loglinear model (3).

Model (3) has a total of 2.k parameters: k-1 random parameters for the k itemresponses, where one parameter is set to zero (i.e. deleted from the model) to fix the scale and k+1 fixed parameters for each of the score groups (t=0,...,k) parametrized, one mean effect u and (k+1)-1 deviation effects $u_1(t)$ for the scores.

Quasi Loglinear Rasch Models for Equating

To use the loglinear Rasch model for equating we have to simultaneously estimate a Rasch model in each of tne equating groups, where the common items have the same parameter over subgroups.

When in each equating group the same total number of items and the same set of common items is used in all equating groups, an equating group x score x response 1 x response 2 x ... x response k contingency table can be formed, where one way is added to the table of the preceding section for the equating group and where each of the items common to all equating groups correspond to one response way. The items that are specific to equating groups correspond to arbitrary remaining response ways, so that these response ways arise from different items in different equating groups. The usual equating design contains the same number of items in the two tests. This design is, for example, found in section preequating. In this paper we will restrict attention to this situation. A method to use the loglinear Rasch model to equate

tests with different numbers of items is briefly discussed at the end of this paper.

As an example, consider a section preequating study where k items are administered to each group. The first r items with responses $X_1,\ldots,X_r$ are common to all equating groups and the remaining k-r items are different in each equating group and have responses $X_{r+(i-1)(k-r)+1}\cdots X_{r+i(k-r)}$ in equating group i. In each equating group the k item responses are denoted by $Y_1,\ldots,Y_k$ regardless of whether the responses comes from different items in different groups or not. In Table 2 an equating design is shown for

---

Insert Table 2 about here

---

two equating groups with two responses (Y_1 and Y_2) from the same items in each group and four responses (Y_3 through Y_6) that come from different items in each group.

To formulate a loglinear Rasch model where the common items have the same item difficulty parameters in both equating groups and the different items have different item difficulty parameters in different equating groups the former responses must be explained by main effects only whereas the latter response must be explained by terms describing an interaction between the equating group and the response. This model is the Rasch equating model:

$$(6) \qquad \ln m_{ity_1 \ldots y_k} = u + u_1(i) + u_2(t) + u_{12}(it) + u_3(y_1) + \ldots + u_{k+2}(y_k) + u_{1(r+3)}(iy_{r+1}) + \ldots + u_{1(k+2)}(iy_k)$$

For $i=1,\ldots,s$; $y_1=0,1;\ldots;y_k=0,1$; $t=y_1+\ldots+y_k$, where $m_{ity_1 \ldots y_k}$ is the conditional expected frequency of the response $Y_1=y_1,\ldots,Y_k = y_k$ for given equating group i and score t.

The u-terms in the model are subject to the constraints

$$u_1(+) = u_2(+) = u_{12}(+t) = u_{12}(i+) = u_3(+) = \ldots = u_{k+2}(+) =$$

$$u_{1(r+3)}(+y_{r+1}) = u_{1(r+3)}(i+) = \ldots = u_{1(k+1)}(+y_k) = u_{1(k+2)}(i+) = 0,$$

where a plus sign indicates summation.

Like in ordinary analysis of variance the u-terms in model (6) describe main and interaction effects of equating group i, score t and responses $y_1,\ldots,y_k$. In model (6) the terms $u_1(i)$, $u_2(t)$, $u_3(y_1),\ldots,u_{k+2}(y_k)$ are the main effects of equating group i, score t, and response $y_1,\ldots,y_k$ respectively. The terms $u_{12}(it)$, $u_{1(r+3)}(iy_{r+1}),\ldots,u_{1(k+2)}(iy_k)$ are the interaction effects of the combination of equating group i and the score t and the combinations of group i and each of the responses $y_{r+1},\ldots,y_k$ respectively.

Model (6) has $s(k+1)+(k-1)+(s-1)(k-r)$ estimable parameters: $s(k+1)$ fixed parameters for the combinations of s equating groups and k+1 scoregroups ($t=0,\ldots,k$), (k-1) random parameters for the

main effects of k item responses, where one item parameter is set to zero to fix the scale, and (s-1)(k-r) interaction parameters for the deviance of (k-r) item parameters in s equating groups from the main effect.

Model (6) is an extension of model (3) for s equating groups. Looking at one equating group only, model (6) has the same linear structure as model (3). Therefore model (3) specifies a Rasch model in each equating group. For example the general mean term u in model (3) corresponds with the sum $u+u_1(i)$ in model (6), the parameter $u_2(x_1)$ corresponds to the parameter $u_3(y_1)$ of the first common item, the parameter $u_{k+1}(x_k)$ corresponds to the sum of the item parameters $u_{k+2}(y_k) + u_{1(k+2)}(iy_k)$ for response of the last noncommon item in equating group i. The item difficulty parameter of tnis noncommon item is equal to $2(u_{k+2}(0)+u_{1(k+2)}(i0))$. Tabel 3 shows wnich sums in model (6) correspond to u terms in model (3).

Insert Taole 3 aoout here

The goodness-of-fit of quasi loglinear models sucn as (6) can oe tested by the usual Pearson statistic

$$Q = \sum_{i} \sum_{y_1} \cdots \sum_{y_k} \frac{(f_{ity_1 \cdots y_k} - m_{ity_1 \cdots y_k})^2}{m_{ity_1 \cdots y_k}}$$

with $t=y_1+\ldots+y_k$ which is asymptotically distributed as chi-square with degrees f freedom equal to the difference between the number $s2^k$ of structurally nonzero cells and the number of estimable parameters.

If Q is significant for the loglinear equating model (5) it means that the items in all subsets of items do not belong to the same Rasch scale.

If there is only one common item, and the parameter of that common item is fixed to zero to fix the scale in both sets of items and to put the item parameter on the same scale estimating model (6) is then equivalent to estimating the loglinear Rasch model (3) for each equating group separately with the parameter of that common item fixed to zero. The function of the common item is merely to use the indeterminacy in both scales to put the item difficulties of both sets of items on the same scale. If there is more than one common item, model (6) is more restrictive than the loglinear Rasch model (3) for each equating group separately with the parameter of one common item fixed to zero. Because in the latter case the parameters of the other common items may vary over equating groups, whereas in model (6) these parameters are the same in different equating groups.

Consequently, if there is more than one common item, the invariance of the parameters of the common items over subgroups is a hypothesis that may or may not be supported by the data. This hypothesis can be tested by comparing model (6) with the model:

$$(7) \qquad \ln m_{ity_1,\ldots,y_k} = u + u_1(i) + u_2(t) + u_{12}(it) + u_3(y_1) + \ldots + u_{k+2}(y_k) + u_{13}(iy_1) + \ldots + u_{1(z+2)}(iy_z) + u_{1(r+3)}(iy_{r+1}) + \ldots + u_{1(k+2)}(iy_k)$$

with the usual constraints.

Model (7) differs from model (5) in that the first z ($1<z<r$) common item parameters may vary over equating groups. Model (7) is the same as model (5) exept for the z interaction terms $u_{13}(iy_1)+\ldots+u_{1(z+2)}(iy_z)$ that describe tne effect of the deviation of common item parameters over subgroups. Model (7) is an extension of model (5) and nas $s.(k+1)+(k-1) + (s-1)(k-(r+z))$ estimable parameters. Its overall goodness of fit can be tested by the Pearson goodness of fit test.

The goodness of fit of model (7) can be compared with the goodness of fit of model (5) by the likelihood-ratio test:

$$G^2 = 2 \sum_{i} \sum_{y_1} \cdots \sum_{y_1} m^{*}_{ity_1\cdots y_k} \ln \left(\frac{m^{*}_{ity_1\cdots y_k}}{m_{ity_1\cdots y_k}} \right)$$

where the starred expected counts are from larger model (7), and the nonstarred expected counts are from the smaller model (5). Under model (5), G^2 is approximately distributed as chi-square with degrees of freedom equal to the difference in numbers of estimable parameters in both models.

Testing model (5) against model (7) using the likelihood-ratio test is a test of the hypothesis that the parameters in which both

models differ are zero given model (7). If both models differ significantly it can be concluded that the parameters of the first z common item parameters are not invariant over subgroups.

The likelihood-ratio test applied to the difference of model (7) and (5) is a test for the goodness of link for equating the scales using the common items. If the test is significant it means that the relative difficulties vary over equating groups. In that case the common items do not belong to the same Rasch scale in each of the equating groups and consequently this link cannot be used to equate these scales.

We have seen before that there must be at least two items to test goodness of link. In this case z is equal to one since one common item parameter must be set to zero to remove the indeterminacy from the model by fixing the scale. If there are more than two common items several subsets of items can be used for testing. If all common items parameters may interact with the equating group model (7) is equivalent to estimating a loglinear Rasch model (3) seperately for each equating group.

If the Rasch equating model (6) fits the data, the usual item difficulty parameters δ_j can be obtained for each of the original items from the item parameters in the model. The common item difficulty parameters δ_j are, as in the loglinear Rasch model (3), equal to twice the model parameter $u_{2+j}(0)$. Whereas for the items that differ in each equating group the item parameters are twice the sum $u_{2+j}(0) + u_{1(2+j)}(i0)$ for item j in group i.

Given these item parameters the testscore of an individual on

any test composed of a subset of the equated items can now be transformed in an ability parameter for that individual (Fischer, 1975; Hambleton and Swaminatan, 1985) by solving the equation:

$$t = \sum_j \exp(\alpha-\delta_j)/[1+\exp(\alpha-\delta_j)]$$

for the ability parameter α. Where the score is obtained from the subset of items used and the item difficulty parameters are obtained from the item parameters in the loglinear model in the way described above. Since the item difficulties are measured on the same scale, the individuals' abilities are also measured on the same scale regardless of the choice of items.

An example

Ten solid geometry items were developed and administered by The Netherlands' Central Institute for Test Development (CITO) to 202 students in secondary education (MAVO). The sample was randomly devided in two halves. In each group four items were removed to form two equating groups with two common and four different items according to the pattern in Table 2. An Equating Group x Score x item 1 x Item 2 x Item 3 x Item 4 x Item 5 x Item 6 contingency table was formed, omitting the structurally zero cells. Both the Rasch equating model (5) and the null model (7) where fitted to the data, using the computer program GLIM (Baker & Nelder, 1978).

Insert Tabel 4 about here

The Rasch equating model fits the data very well (Q = 111.1, df=102) and the model does not differ significantly from the null model (G^2 = 0.0, df=1). The estimates and standard errors of the item parameters are shown in Table 4. Taking twice the sum of the main effect parameter and the interaction effect parameter as described above, for each of the original items x (see Table 2) the item difficulty parameters are calculated. These item difficulty parameters are given in Table 5.

Insert Table 5 about here

Discussion

In this paper a loglinear-Rasch model is formulated that can be used for equating. An equating group x score x response 1 x ... x response k contingency table was formed where a number of responses are from different items in different equating groups. This approach can be used if the number of items to be analysed is the same in each equating group and if the same common items are used. Most equation designs are within these restrictions.

If a particular equating design does not stay within these

restrictions, the above contingency table approach is no longer feasible and a designmatrix approach must be used. In the design matrix approach (Evers & Namboodiri 1969; Imrey, Koch, & Stoker, 1981; MacGullagh & Nelder 1983) each parameter corresponds to a column in the designmatrix. The model can now be estimated by giving all the items a column in the design matrix: for the equating groups where an item is not used the entries of its column are given the value zero, but for the equating groups where the item is used, the column is given the usual contrast codes (Bock 1975). In this case parameter estimation and hypothesis testing can also be done with the GLIM computer program.

In this paper we work with an incomplete tabel having 2^k structurally nonzero cells. If the number of items is large the table becomes very large and many cells will be empty. Most present day computer programs for the analysis of contingency tables by loglinear models (Baker & Nelder, 1978; Goodman & Fay, 1974) require the internal storage of the observed and expected table of couts, which is virtually impossible if the number of items is large. In principle, however, this is unnencessary. the table need only exist in theory. first, the data can be stored in an ordinary N x k data matrix, which avoids storage of structural and nonstructural zero's. Second, the parameter estimates may be calculated by solving the likelinood equations in terms of the minimal set of sufficient marginal tables. Kelderman (1986) describes an efficient algorithm for the calculation of the parameters of quasi-loglinear models for the incomplete Rasch table

that uses such methods.

Another problem in the analysis of large Rasch tables concerns the approximation of the overall goodnes of fit statistics to the chi square distribution. If the expected cell counts become small, this approximation is known to be bad (Lancaster, 1961). In this case the fit of the Rasch equating model can be tested with the lielihood-ratio test against a larger Rasch equating model that contains the parameters of the Rasch model but also parameters describing deviations of the Rasch equating model. As deviations we can take, for example, all interactions of order, say, order four.

The general loglinear model is a very general framework within which Rasch equating data can be analyzed. In theory, there are no limits to the number of categorical variables that can be analyzed. Consequently the models discussed in this paper can be extended with parameters describing the effects of one or more backgroud variables; e.g. sex, race; on item difficulty to study possible item bias (Kelderman, 1985).

Furthermore the models may be extended with more sumscore variables so that several Rasch equating scales may be analyzed simultaneously.

References

Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington, D.C.: Americal Council on Education, 508-600.

Baker, R.J. & Nelder, J.A. (1978). The GLIM system: Generalized linear interactive modelling. Oxford: The Numerical Algorithms Group.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). Discrete multivariate analysis. Cambridge, Mass.: MIT Press.

Bock, R.D. (1975). Multivariate Statistical Methods in Behavioral research. New York: McGraw Hill.

Divgi, D.R. (1981). Model-free evaluation of equating and scaling. Applied Psychological Measurement 5, 203-208.

Evers, M. & Namboodiri, N.K. (1979). On the design strategy in the analysis of categorical data. In Karl F. Schuessler (Ed.) Social Methodology. San Francisco: Jossey-Bass.

Fienberg, S.E. (1972). The analysis of incomplete multi-way contingency tables. Biometrics, 28, 177-202. Corrig. 1972, 29, 829.

Fischer, G.H. (1974). Einfuenrung in die Theorie psychologischer Tests. Bern: Verlag Hans Huber.

Guskey, T.R. (1981). Comparison of a Rasch model scale and the Grade Equivalent Scale for the vertical equating of test scores. Applied Psychological Measurement, 5, 187-201.

Gustafsson, J.E. (1979). Rasch Model in vertivcal equating of tests - critique of Slinde and Linn. Journal of Eductional Measurement, 16, 153-158.

Hambleton, R.K. & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston: Kluwer Nijhoff Publishing.

Imrey, P.B., Kock, G.C. and Stoker, M.E. (1981). Categorical data analysis: Some reflections on the log-linear model and logistic regression. Part I: Historical and methodological overview. International Statistical Review, 49, 265-283.

Kelderman, H. (1984). Loglinear Rasch Model Tests Psychometrika, 49, 223-245.

Kelderman, H. (1985). Item bias detection using the loglinear Rasch model: observed and unobserved subgroups. Paper presented at the Fifthieth annual meeting of the Psychometric Society, Nashville, Tennesse, june 1-4, 1985.

Lancaster, H.O. (1961). Significance tests in discrete distributions. Journal of the Americal Statistical Association, 56, 223-234.

Lord, F.M. & Novick, M.R. (1968). Statistical tneories of mental test scores. Reading, Mass.: Addison-Wesley.

Lord, F.M. (1977). Practical applications of item characteristic curve tneory. Journal of Educational Measurement, 14, 117-138.

Loyd, B.H., & Hoover, H.D. (1980). Vertical equating using the Rasch model. Journal of Educational Measurement, 17, 179-193.

Masters, G.N. (1985). Common person equating with the Rasch model. Applied Psychological Measurement, 9, 73-82.

McCullagh, P.M. & Nelder J.A. (1983). Generalized Linear Models. London: Chapman and Hall.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Paedagogiske Institut.

Rasch, G. (1966). An item analysis that takes individual differences into account. Britisn Journal of Mathematical and Statistical Psychology, 19, 49-57.

Rentz, R.R. & Bashaw, W.L. (1975). Equating Reading with the Rasch Model (Vols. 1 and 2). Athens: Educational Research Laboratory, College of Education, University of Georgia.

Slinde, J.A. & Linn, R.L. (1977). Vertically equated tests: Fact or phantom? Journal of Eductional Measurement, 14, 23-32.

Slinde, J.A. & Linn, R.L. (1978). An exploration of the adequacy of the Rasch model for the problem of vertical equating. Journal of Educational Measurement, 15, 23-35.

Slinde, J.A. & Linn, R.L. (1979). Vertical equating via the Rasch model for groups of quite different ability and tests of quite different difficulty. Journal of Educational Measurement, 16, 159-165.

Wright, B.D. (1977). Solving measurement problems with the Rasch model. Journal of Educational Measurement, 14, 97-116.

Table 1

Frequency Counts and Structural Zero's in Score x Item 1 x ... x Item 3 Table.

Item Response			Score t			
			0	1	2	3
x_1	x_2	x_3				
0	0	0	f_{0000}	-	-	-
1	0	0	-	f_{1100}	-	-
0	1	0	f_{1010}	-	-	-
0	0	1	-	f_{1001}	-	-
1	1	0	-	-	f_{2110}	-
1	0	1	-	-	f_{2101}	-
0	1	1	-	-	f_{2011}	-
1	1	1	-	-	-	f_{3111}

Note: Dashes denote structurally zero cells.

Table 2

Item Responses in Equating Groups Corresponding to Response Ways in a Rasch Equating Table

Equating Group	Response Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
1	X_1	X_2	X_3	X_4	X_5	X_6
2	X_1	X_2	X_7	X_8	X_9	X_{10}

Table 3

Correspondence of u-terms between model (3) and (6)

Model (3)	Model (6)
u	$u + u_1(i)$
$u_1(t)$	$u_2(t) + u_{12}(it)$
$u_2(x_1)$	$u_3(y_1)$
. . . .	
$u_{r+1}(x_r)$	$u_{r+2}(y_r)$
. . . .	
$u_{k+1}(x_k)$	$u_{k+2}(y_k) + u_{1(k+2)}(iy_k)$

Table 4

Item Parameter Estimates (and Standard Errors) in the Rasch Equating model for the Solid Geometry Data

	Response variable j					
Parameter	1	2	3	4	5	6
$u_{2+j}(0)$	0.00*	0.80	-0.29	2.268	1.49	2.09
	(-)	(.12)	(.24)	(.27)	(.24)	(.2o)
$u_{1(2+j)}(10)$	0.00*	0.00*	-0.22	0.77	0.55	0.80
	(-)	(-)	(.10)	(.11)	(.11)	(.12)

Note: parameters with an asterix are set to zero.

Table 5

Difficulty Parameters for the Original Solid Geometry Items

Item									
1	2	3	4	5	6	7	8	9	10
0.00*	1.60	-1.02	6.07	4.08	5.78	-0.14	3.00	1.88	2.58

Note: parameters with an asterix are set to zero.

Titles of Recent Research Reports

RR-86-1 W.J. van der Linden, The use of test scores for classification decisions with threshold utility

RR-86-2 H. Kelderman, Item bias detection using the loglinear Rasch model: Observed and unobserved subgroups

RR-86-3 E. Boekkooi-Timminga, Simultaneous test construction by zero-one programming

RR-86-4 W.J. van der Linden, & E. Boekkooi-Timminga, A zero-one programming approach to Gulliksen's random matched subtests method

RR-86-5 E. van der Burg, J. de Leeuw, & R. Verdegaal, Homogeneity analysis with k sets of variables: An altenating least squares method with optimal scaling features

RR-86-6 W.J. van der Linden, & T.J.H.M. Eggen, An empirical Bayes approach to item banking

RR-86-7 E. Boekkooi-Timminga, Algorithms for the construction of parallel tests by zero-one programming

RR-86-8 T.J.H.M. Eggen, & W.J. van der Linden, The use of models for paired comparisons with ties

RR-86-9 H. Kelderman, Common Item Equating Using the Loglinear Rasch Model

RR-86-10 W.J. van der Linden, & M.A. Zwarts, Some Procedures for Computerized Ability Testing

Research Reports can be obtained at costs from Mediatheek, Faculteit Toegepaste Onderwijskunde, Universiteit Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

A publication by
the Department of Education
of the University of Twente,
P.O. Box 217,
7500 AE Enschede,
The Netherlands

Department of Education